R E P O R T R E S U M E S

ED 015 006 PS 000 205

DEVELOPMENT OF APPROPRIATE EVALUATION TECHNIQUES FOR SCREENING CHILDREN IN A HEAD START PROGRAM. A PILOT PROJECT.

BY- BERGER, STANLEY I.

REPORT NUMBER OEO-515
EDRS PRICE MF-$0.25 HC-$0.60 13P.

DESCRIPTORS- CULTURAL DISADVANTAGEMENT, INTELLIGENCE TESTS, *TEST VALIDITY, TEST SELECTION, *EVALUATION TECHNIQUES, *SCREENING TESTS, PROGRAM EVALUATION, PRESCHOOL TESTS, INTELLECTUAL DEVELOPMENT, VERBAL ABILITY, *EARLY EXPERIENCE, TEST RELIABILITY, *COGNITIVE MEASUREMENT, MEASUREMENT INSTRUMENTS, HEAD START, PPVT, LEITER INTERNATIONAL, RAVEN PROGRESSIVE MATRICES, STANFORD BINET,

THE PURPOSES OF THIS PILOT PROJECT WERE (1) TO ATTEMPT TO EVALUATE THE EFFECT OF THE LOCAL PROGRAM ON BOTH INDIVIDUAL CHILDREN AND THE GROUP AND (2) TO INVESTIGATE THE SENSITIVITY OF THE TEST INSTRUMENTS EMPLOYED IN EVALUATING SUCH A PROGRAM. SIXTY-ONE CHILDREN WERE ENROLLED IN THE LOCAL HEADSTART PROGRAM AND WERE ADMINISTERED THE STANFORD-BINET, LEITER INTERNATIONAL, RAVEN PROGRESSIVE MATRICES, AND PEABODY PICTURE VOCABULARY TESTS. IN ADDITION, 20 CHILDREN, SELECTED AT RANDOM FROM THE GROUP WERE TESTED BOTH BEFORE AND AFTER THE PROGRAM. RESULTS INDICATE (1) STATISTICALLY SIGNIFICANT IMPROVEMENT IN PERFORMANCE FOR THE 20 CHILDREN, (2) SIGNIFICANT CORRELATIONS AMONG THE VARIOUS TEST SCORES OF THE TOTAL GROUP, AND (3) PARTICULAR SENSITIVITY OF THE LEITER AND PEABODY TESTS IN REFLECTING CHANGES IN FUNCTIONING. IMPLICATIONS OF THE STUDY FOR FUTURE HEADSTART PROGRAMS AND ALSO FOR FURTHER RESEARCH WITH CULTURALLY DEPRIVED CHILDREN WERE DISCUSSED. (CO'D)

ED015006

2459

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

OEO-515

PROJECT HEADSTART

Development of Appropriate Evaluation Techniques for Screening Children in a Head Start Program - A Pilot Project

PS 000205

Dr. Stanley I. Berger

Project Coordinator

2460

ACKNOWLEDGEMENTS

I wish to express my gratitude to the Office of Economic Opportunity for making this project possible and to the personnel and teachers of Head-Start, especially Miss Sylvia Lapin, whose support for and interest in the project proved of invaluable assistance. My special appreciation goes to Mr. Neal Dye for his specific assistance in analysis of the data.

Dr. Stanley I. Berger

2461

ABSTRACT

Sixty-one children enrolled in the local Headstart program and were administered the Stanford-Binet, Leiter International, Raven Progressive Matrices, and Peabody Picture Vocabulary tests. In addition, 20 Ss selected at random from the group were tested both before and after the program. Results indicate: (1) statistically significant improvement in performance for the 20 Ss, (2) significant correlations among the various test scores of the total group, and (3) particular sensitivity of the Leiter and Peabody tests in reflecting changes in functioning.

Implications of the study for future Headstart programs, and also for further research with culturally deprived children, were discussed.

2462

Introduction

Since the early Iowa studies of the forties (Wellman, 1940, 1945) there has been an increasing amount of literature suggesting that early pre-school experience has a profound impact upon the subsequent intellectual development of the child. An evaluation of such a program must be concerned with the reliability and validity of the instruments employed. Project Headstart thus creates several problems for the investigator concerned with measuring the effect of the program on the culturally deprived child. One such issue concerns the development of criteria to be used for selection of truly "culturally deprived" children; i.e., the selection of a homogenous group with regard to this variable. A second variable is related to the nature of the instruments utilized in evaluating the effects of the program on both individual and the group.

The cultural factor involved in most intelligence instruments is rather well established (Cronbach, 1960). Instruments such as the Stanford-Binet, which primarily require verbal ability, illustrate the effect of cultural differences most clearly (Davis, 1951; Havighurst and Janke, 1944, 1945; Eells, 1951; Thurstone, 1951). This poses something of a dilemma in evaluating the Headstart program, since the Binet is also recognized as the best single predictor of scholastic readiness available (Cronbach, 1960). Thus in developing a battery for this group it would seem necessary to include instruments in which items are less heavily weighted for verbal ability, but which yield reliable estimates of intellectual potential. That is, instruments should be included which essentially correlate with Binet IQ's.

In light of the apparently tenable assumption that cultural deprivation would vary both quantitatively and qualitatively with the particular geographic area, it seems further appropriate that some evaluation of the instruments be carried out for a given locale. Not only would this provide some indication of their reliability and validity for this group, but it might also yield information regarding which

2463

instrument might optimally be employed to reflect their unique experiences. That is it is desirable to ascertain which instrument provides an optimal estimate of potential intellectual ability of the deprived child.

Thus the purposes of this pilot project were:

1) an attempt to evaluate the effect of the local program on both individual children as well as the group,
2) an investigation of the test instruments employed in order to derive information regarding the sensitivity of such media in evaluating such a program.

Method

Subjects. Sixty-one, five-year-old children of essentially rural, low-income background, who were enrolled in the program.

Instruments. The traditional instrument yielding IQ scores employed was the Stanford-Binet (1960, L-M). Included also were the Leiter International, Raven Progressive Matrices, and Peabody Picture Vocabulary tests, for essentially non-verbal indices of ability. A "pre-school inventory" was also given to most children.

Procedure. All Ss were administered at least some of the tests in a period between one week prior to the start of the program - which lasted 7 weeks - through the end of the last week. In addition, 20 Ss were randomly selected from the group for pre- and post-evaluations. These were given the above tests within one week of the start of the program, and again during the last week and one week thereafter. As a preliminary pilot study it was decided that the limited time available required total attention and activity be directed toward this group, and thus a control group could not be included.

2464

Results and Discussion

A. Evaluation of total group. As scoring norms and standards were not available for the pre-school inventory no evaluation of the "P.I." was attempted.

Means and standard deviations were computed for scores on each test (see Table 1 below).

TABLE 1

MEANS, STANDARD DEVIATIONS, AND RANGES FOR

SCORES OF EACH INSTRUMENT

	N	$\bar{X}$	σx	Range
Binet	53	91.62	11.65	56-116
Leiter	59	84.20	10.88	50-111
Raven	61	4.09	1.73	0-8
Peabody	59	82.80	12.04	35-110

It should immediately be noted that the obtained scores (means and sigmas) are inconsistent with what one would logically expect for a truly culturally deprived group. The extent to which the group deviates from expected poor scores would limit in itself the application of our findings as regards the culturally deprived child. Thus, the inclusion of children not culturally deprived could have an influence on all statistical analysis and interpretation, and all conclusions must be qualified with this in mind. In light of the lack of objective individual correlates of cultural deprivation, however, it is impossible to delete individual Ss from the analysis purely on the basis of high intelligence test scores or cultural background.

B. Evaluation of the instruments. Pearson Product-Moment Correlation Coefficients were computed among scores on all tests (see Table 2 on the following page).

2465

TABLE 2

PRODUCT-MOMENT INTER-CORRELATIONS AMONG SCORES ON EACH TEST

(Ns in parens)

	Binet	Leiter	Raven	Peabody
Binet		.625*(53)	.554*(53)	.510*(53)
Leiter			.693*(59)	.437 (59)
Raven				.412 (59)

* significant at p = .05

Although not extremely high, and despite a large standard error (.23), these data do indicate some degree of overlap. If the Binet can be accepted as being one of the more reliable available tests, then it would appear that, for this group, a reasonable estimate of intellectual potential can be obtained by employing other instruments. Being primarily interested in obtaining an optimal estimate of functioning, these data would seem to justify the use of the Leiter and Peabody tests in a program of this sort, despite flaws in standardization data. This would become especially significant where a child indicates difficulty with verbal items.

C. Evaluation of the pre- and post-program group. Matched-group t-tests computed between pre- and post-program test scores indicate statistically significant improvement in all test scores except the Raven (see Table 3 on the following page).

The pragmatic or clinical significance, however, of a 3.4 point gain (as in the Binet) is questionable, especially since it is well within the standard error of measurement.

3466

TABLE 3

t-TESTS BETWEEN PRE- AND POST-PROGRAM TEST SCORES

(N=20)

Instrument	d	$t_{1.19}$	p*
Binet	+3.4	1.61	.01
Leiter	+10.55	3.47	.001
Raven	-0.35	0.778	.05
Peabody	+9.75	3.61	.001

* one-tailed tests of significance

In order to more specifically assess any changes in performance, Binet items were separated into those requiring primarily verbal, performance, and memory ability (McNemar, 1942), and t-tests were conducted on the proportion of items passed in each category. No significant differences were found either between pre- and post-program scores for each category, or among the categories when measured as units.

In an attempt to compare the various scores for each individual, the group was ranked on each test according to their standardized score position relative to the group, for both the pre- and post-programs. Kendall Coefficients of Concordance - "W" - (Siegel, 1956) computed among ranks indicated no significant changes in rank within the group or among tests, either pre- or post-program. That is, Ss tended to do as well or as poorly - relative to the group - on each test, both before and after their Headstart experience. It would appear, then, that improvement was approximately equal for most Ss, and that no particular instrument was easier or more difficult for the group or for the individual Ss.

It should be noted, however, that a very few individuals did respond uniquely, particularly in an "upward" direction. The small number involved precludes statistical analysis.

It is impossible, in a study of this sort, to separate the effects of practice from those of true improvement from the Headstart experience. Also, questions about reliability of all the instruments for this age group necessitates caution in interpreting these results. It cannot, therefore, be stated at this time whether the improvement in scores is specifically due to the beneficial effects of Headstart, although one would suspect that the total gain in scores would not likely be due solely to practice, especially for such an age group.

D. "Free association" comments made by psychometricians reflecting upon their testing experience with the children. Some of the children refused to cooperate and many of these were very low on the testing that was done. Hence there is a possibility that this sample obtained may not be representative, since many of the lower cases were not included. There did not seem to be any noticeable difference between male and female children. For some of the children there was a noticeably short attention span. They were quite hyperactive, couldn't sit through one test. Some had language handicaps and some were almost unintelligible. Some were very shy and withdrawn, especially during the pretesting period. D.H. felt that many of the children did do their best, did reach their potential and the testing was an adequate reflection of their abilities. A.W. felt that potentials were never tapped on the test because of the age level of the children and also because of the tests themselves. Maturation level and motor ability affected some of the drawing tasks. Concerning the Stanford-Binet itself, all testers agreed that this test could not be given in one session. When given in one session, just at the time the test began to discriminate at the higher levels, the children almost always became tired and would not cooperate or would give arbitrary answers.

D.H. felt that the picture completion proved to be one of the better tests. The elaboration of the drawing usually correlated with the overall ability of the children. A.W., however, felt that this test was inadequate because of motivational

2468

factors. One of the children in this test attempted a very bizarre drawing with four claws in place of a left arm. Later it was discovered that she was innoculated previously with a four-prong needle which slipped into her arm, hence explaining the bizarre drawing. In this sense, many of the reactions and responses were due to very practical reasons, such as lack of sleep, poor diet, etc., in some of the children. However, a few of the children did produce definitely bizarre responses which were apparent through all the total testing indicating a need for clinical referral.

In summary, verbal ability was below the performance. In this regard, post testing showed improvement in this area. However, almost 7 out of 20 of the sample cases showed a negative trend, with post testing with the Stanford-Binet. Conditions for testing were far from optimal. Children were taken away from playing games and snack time which they were enjoying, adding to the poor motivation previously mentioned. In many cases, testing was done in spite of the teacher's help rather than with the teacher's help. Occasionally they would cue children as to how to respond by saying, "This won't hurt a bit. Don't worry, he's not going to hurt you." This was probably due to the lack of structure in the program and the personnel not knowin really what was expected of them.

Summary and Conclusions

Results of this study indicate the following:

(1) The group may not have been truly homogenous with respect to cultural deprivation.

(2) A significant correlation exists among non-verbal instruments and the Stanford-Binet.

(3) Tests which are primarily oriented towards evolving "performance" rather than verbal behavior appear to be more sensitive in reflecting change as a result of the childrens' becoming more familiar with a benign "action" atmosphere.

(4) A significant gain is achieved in test scores after completion of the Headstart program experience. It should be emphasized, however, that the limitation of the present design does not permit one to assess how much of the improvement might be due to practice effect.

In general, the results of the study indicate that such a program has a probable significant impact upon releasing cognitive attitudes and "sets" which are necessary precursor mechanisms in the learning process. Thus, one sees striking change in test scores, particularly in those tests which facilitate expression of ability in pathways most consistent with the child's own cognitive style. While one anticipates a relative constancy of IQ score, it clearly becomes necessary - particularly with such a group - to have available instruments which are maximally sensitive in reflecting actual as well as potential levels of performance and changes in functioning. While it is possible that the obtained changes may be a function of greater or less test reliability, it is tenable to hold the view that some tests are actually more sensitive than others in reflecting change. Thus the Leiter in particular, as well as the Peabody Picture-Vocabulary, seem to be tests which reflect an ability to pick up both gross as well as subtle shifts in the child's reaction to structured cognitive tasks. If this position can be borne out by further, more rigorous study, it would be important to carefully pre-test any instruments used in evaluating such a program as Headstart in order to determine the degree to which they might mirror actual changes occurring in the children.

The improvement beyond chance expectation of some few Ss would seem to suggest the possibility of differential readiness for such a program as Headstart. This is particularly evident in instruments such as the Leiter and the Peabody, where less emphasis is placed upon strictly verbal ability, and intellectual potential and weaknesses may be more clearly and readily reflected. It would appear tenable to assume that some specific cognitive and personality correlates of cultural deprivatio do exist, and that these can be measured on a properly controlled and rigorously designed longitudinal study. The specific question one might ask would be concerned

2470

with the prediction of which children might benefit most, or least, from a Headstart experience. This question could only be answered by including besides cognitive instruments, some projective techniques and clinical interviews. A longitudinal study would also allow for the inclusion of a number of external criteria, such as school grades, attitudes, etc.

In conclusion, within the very real limits imposed upon the study by virtue of the lack of adequate controls, the question of homogeneity of the group regarding "cultural" deprivation, and the necessity to maintain time schedules as a critical factor, these data do support the assumption that the experience of the Headstart program can produce effective, positive results.

2471
v.14

REFERENCES

Cronbach, L.J. Essentials of Psychological Testing (2nd ed.), New York: Harper, 1960.

Davis, A. Socioeconomic influences upon childrens' learning. Understanding the Child, 1951, 20, 10-16.

Kells, K., et al. Intelligence and Cultural Differences, Chicago: Univ. of Chicago Press, 1951.

Havighurst, R.J. and L.L. Janke. Relations between ability and social status in a midwestern community. J. educ. Psychol., 1944, 35, 357-358; 1945, 36 499-509.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill, 1956.

McNemar, Q. Revision of the Stanford-Binet Scale, Boston: Houghton-Mifflin, 1942.

Thurstone, L.L. Creative talent. Proceedings, 1950 Invitation Conference on Testing Problems, Princeton: E.I.S., 1951.

Wellman, L.L. IQ changes in pre-school and non-pre-school groups during the pre-school years: a summary of the literature, J. Psychol., 1945, 20, 347-368.